

# Diversity, disparity, and evolutionary rate estimation for unresolved Yule trees

Forrest W. Crawford and Marc A. Suchard

Typeset July 23, 2012

## Abstract

The branching structure of biological evolution confers statistical dependencies on phenotypic trait values in related organisms. For this reason, comparative macroevolutionary studies usually begin with an inferred phylogeny that describes the evolutionary relationships of the organisms of interest. The probability of the observed trait data can be computed by assuming a model for trait evolution, such as Brownian motion, over the branches of this fixed tree. However, the phylogenetic tree itself contributes statistical uncertainty to estimates of other evolutionary quantities, and many comparative evolutionary biologists regard the tree as a nuisance parameter. In this paper, we present a framework for analytically integrating over unknown phylogenetic trees in comparative evolutionary studies by assuming that the tree arises from a continuous-time Markov branching model called the Yule process. To do this, we derive a closed-form expression for the distribution of phylogenetic diversity, which is the sum of branch lengths connecting a set of taxa. We then present a generalization of phylogenetic diversity which is equivalent to the expected trait disparity in a set of taxa whose evolutionary relationships are generated by a Yule process and whose traits evolve by Brownian motion. We derive expressions for the distribution of expected trait disparity under a Yule tree. Given one or more observations of trait disparity in a clade, we perform fast likelihood-based estimation of the Brownian variance for unresolved clades. Our method does not require simulation or a fixed phylogenetic tree. We conclude with a brief example illustrating Brownian rate estimation for thirteen families in the Mammalian order Carnivora, in which the phylogenetic tree for each family is unresolved.

**Keywords:** Brownian motion, Comparative method, Markov reward process, Phylogenetic diversity, Pure-birth process, Quantitative trait evolution, Trait disparity, Yule process

# 1 Introduction

Evolutionary relationships between organisms induce statistical dependencies in their phenotypic traits (Felsenstein, 1985). Closely related species that have been evolving separately for only a short time will generally have similar trait values, and species whose most recent common ancestor is more distant will often have dissimilar trait values (Harvey and Pagel, 1991). However, the origins of phenotypic diversity are still poorly understood (Eldredge and Gould, 1972; Gould and Eldredge, 1977; Ricklefs, 2006; Bokma, 2010). Even simple idealized models of evolutionary change can give rise to highly varying phenotype values (Foote, 1993; Sidlauskas, 2007), and researchers disagree about the relative importance of time, the rate of speciation, and the rate of phenotypic evolution in generating phenotypic diversity (Ricklefs, 2004; Purvis, 2004; Ricklefs, 2006).

Comparative phylogenetic studies seek to explain phenotypic differences between groups of taxa, and stochastic models of evolutionary change have assisted in this task. Researchers often treat phenotypic evolution as a Brownian motion process occurring independently along the branches of a *fixed* macroevolutionary tree (Felsenstein, 1985). In comparative studies, the Brownian motion model of trait evolution has a convenient consequence: given an evolutionary tree topology and branching times, the trait values at the concurrently observed tips of the tree are distributed according to a multivariate normal random variable. Brownian motion on a fixed phylogenetic tree is the basis for the most popular regression-based methods for comparative inference and hypothesis testing (Grafen, 1989; Garland et al, 1992; Martins and Hansen, 1997; Blomberg et al, 2003; O’Meara et al, 2006; Revell, 2010). In the regression approach, inference of evolutionary parameters of interest becomes a two-step process: first, one must infer a phylogenetic tree; then, *conditional on that tree*, one estimates relevant evolutionary parameters, usually by maximizing likelihood of the observed trait data under the model for trait evolution. Unfortunately, the uncertainty involved in estimating the tree propagates into the comparative analysis in a way that is difficult to account for (but see Stone (2011)), and comparative researchers often lack a precise phylogenetic tree on which to base a regression analysis of trait data. Modern techniques for dealing with this issue generally resort to simulation. Some researchers simulate a large number of possible trees and estimate parameters conditional on a single representative tree, such as the maximum clade credibility tree (see, for example, Alfaro et al (2009)). Alternative approaches that utilize simultaneous simulation

of trees and parameters via Bayesian methods are gaining in popularity (Sidlauskas, 2007; Slater et al, 2012; Drummond et al, 2012).

However, simulation methods can be extremely slow and may require assumptions about prior distributions of unknown parameters that are difficult to justify. Indeed, in macroevolutionary studies, the phylogenetic tree is often not of interest *per se*, but must be taken into account in order to accurately model the dependency of the traits under consideration. Many comparative phylogeneticists regard the evolutionary tree as a nuisance parameter in the larger evolutionary statistical model. For this reason, there is increased interest in tree-free methods of comparative analysis that preserve information about the variance of phenotypic values within unresolved clades (Bokma, 2010).

To develop a method for comparative inference in evolutionary studies that does not rely on a particular tree, it is convenient to specify a *generative* model for phylogenetic trees. In the Yule (pure-birth) process, every existing species independently gives birth with instantaneous rate  $\lambda$ ; when there are  $n$  species, the total rate of speciation is  $n\lambda$  (Yule, 1925). The Yule process is widely used as a null model in evolutionary hypothesis testing and can provide a plausible prior distribution on the space of evolutionary trees in Bayesian phylogenetic inference (Nee et al, 1994; Rannala and Yang, 1996; Nee, 2006). One can easily derive finite-time transition probabilities (Bailey, 1964), and efficient methods exist to simulate samples from the distribution of Yule trees, conditional on tree age, number of species, or both (Stadler, 2011). Interestingly, some researchers have pointed out that even the simple Yule process can have unexpected properties that may be relevant in evolutionary theory and reconstruction (Gernhard et al, 2008; Steel and Mooers, 2010).

Due to Yule trees' simple Markov branching structure and analytically tractable transition probabilities, many researchers have made progress in characterizing summary properties of the Yule process – that is, integrating over all Yule tree realizations. For example, Steel and McKenzie (2002) study aspects of the shape of phylogenies under the Yule model, such as the distribution of the number of edges separating a subset of the extant taxa from the MRCA; Gernhard et al (2008) find distributions of branch lengths; Steel and Mooers (2010) study the expected length of pendant and interior edges of Yule trees; and Steel and McKenzie (2001) and Mulder (2011) study the distribution of the number of internal nodes separating taxa.

One important summary statistic for trees in biodiversity applications is *phylogenetic diversity*

(PD), defined as the sum of all branch lengths in the minimum spanning tree connecting a set of taxa (Faith, 1992). Applied researchers in evolutionary biology have found PD to be useful in conservation and biodiversity applications; see, e.g., Webb et al (2002), Moritz (2002), and Turnbaugh et al (2008). PD also has the virtue of being a mathematically tractable statistic for phylogenetic trees, and has attracted interest from researchers interested in its properties. For example, Faller et al (2008) show that the asymptotic distribution (as the number of taxa  $n \rightarrow \infty$ ) of PD is normal and give a recursion for computing the distribution of PD where edge lengths are integral. Mooers et al (2011) discuss branch lengths on Yule trees and expected loss of PD in conservation applications. Most importantly for our study, Stadler and Steel (2012) find the moment-generating function for PD conditional on  $n$  extant taxa and tree age  $t$  under the Yule model. Following on these inspiring results, we seek now to study analytic properties of Yule trees that are useful for comparative evolutionary studies.

In this paper, we present a framework for computing probability distributions related to diversity and quantitative trait evolution over unresolved Yule trees and describe methods for estimating related parameters. We first give a mathematical description of the Yule model of speciation and briefly discuss its properties. Next, we introduce the Markov reward process, a probabilistic method for deriving probability distributions related to the accumulation of diversity under a Yule model. In Theorem 1, we give an expression for the probability distribution of PD under a Yule model, conditional on the number of species  $n$ , time to the most recent common ancestor (TMRCA) or tree age  $t$ , and speciation rate  $\lambda$ . We then demonstrate an important and previously unappreciated relationship between *trait disparity*, the sample variance for a group of taxa (O’Meara et al, 2006), and PD for traits evolving on a Yule tree via Brownian motion. Theorem 2 gives an expression for the distribution of expected trait disparity when integrating over the branch lengths of a Yule tree. Next, we describe a statistical method for performing fast maximum likelihood estimation of Brownian variance, given an unresolved clade and observed trait disparity. Our approach does not require fixing a phylogenetic tree or specification of prior probabilities for unknown parameters. The method is simulation-free and does not seek to infer branch lengths or ancestral states. We show empirically that our estimators are asymptotically consistent. We conclude with an application of our method to body size evolution in the Mammalian order Carnivora.

## 2 Mathematical background

To aid in exposition, we briefly establish some notation. Denote the *topology* of a phylogenetic tree by  $\tau$ . A topology is the shape of a tree, disregarding branch lengths or age. We always condition our calculations on the phylogenetic tree having age  $t$  with  $n$  extant taxa. Let  $t_1, t_2, \dots$  denote the branching points of a tree, where  $t_k$  is the time of branching from  $k$  to  $k + 1$  lineages. We measure time in the forward direction, so at the TMRCA,  $t = 0$ . This is in keeping with our mechanistic orientation: the Yule process, to be developed below, runs forward in time from 0 to  $t$ .

### 2.1 Yule processes

Let  $Y(t) \in \{1, 2, \dots\}$  be a Yule process with birth rate  $\lambda$  that keeps track of the number of species at time  $t$ . The transition probabilities  $P_{mn}(t) = \Pr(Y(t) = n \mid Y(0) = m)$  satisfy the Kolmogorov forward equations

$$\begin{aligned} \frac{dP_{m1}(t)}{dt} &= -\lambda P_{m1}(t), \quad \text{and} \\ \frac{dP_{mn}(t)}{dt} &= -\lambda n P_{mn}(t) + (n-1)\lambda P_{m,n-1}(t) \end{aligned} \tag{1}$$

for  $n \geq 1$ . This infinite system of ordinary differential equations can be solved to yield closed forms for the finite-time transition probabilities,

$$P_{mn}(t) = \binom{n-1}{m-1} e^{-m\lambda t} (1 - e^{-\lambda t})^{n-m}, \tag{2}$$

which have a negative binomial form (Bailey, 1964). In the Yule process, we are only concerned with the number of species that exist at any moment in time, not their genealogy. That is, we assume that the lineage that branches is chosen uniformly from all extant lineages. The transition probability (2) is useful for performing statistical inference: suppose we know the branching rate  $\lambda$  and the age  $t$  of a tree, and we observe  $Y(t) = n$ . Then (2) gives the *likelihood* of our observation. Figure 1 shows an example realization of a Yule tree, with the corresponding counting process diagram below. In this example,  $\lambda = 2$ ,  $Y(0) = 2$ , and  $Y(t = 1) = 12$ .

[Figure 1 about here.]

## 2.2 Markov reward processes

In a Markov reward process, a non-negative reward  $a_k$  accrues for each unit of time a Markov process spends in state  $k$  (Neuts, 1995). Consider a Yule process  $Y(s)$  beginning at  $Y(0) = 1$  and ending at  $Y(t) = n$ . The accumulated reward up to time  $t$  is

$$R_t = \int_0^t a_{Y(s)} \, ds. \quad (3)$$

When  $Y(s)$  is observed continuously from time 0 to  $t$ , the process  $a_{Y(s)}$  is a fully-observed step function, and  $R_t$  can be easily computed as the area under that function. To illustrate, suppose that the process makes jumps at times  $t_1, \dots, t_{n-1}$ , and we define  $t_0 = 0$  and  $t_n = t$ . We assume  $Y(s)$  is right-continuous, so  $Y(t_i) = i + 1$ . Then at time  $t$ , the accumulated reward is

$$R_t = \sum_{i=1}^n a_{Y(t_{i-1})}(t_i - t_{i-1}) = \sum_{i=1}^n a_i(t_i - t_{i-1}). \quad (4)$$

When only  $Y(0)$  and  $Y(t)$  are observed, it can be challenging to compute the distribution of  $R_t$ . In our proofs of Theorems 1 and 2, we appeal to the method developed by Neuts (1995) and Minin and Suchard (2008) to find reward probabilities conditional on  $Y(0)$  and  $Y(t)$ . Let

$$v_{mn}(x, t) = \Pr(R_t = x, Y(t) = n \mid Y(0) = m) \quad (5)$$

be the joint probability that the reward at time  $t$  is  $x$  and the process is in state  $n$ , given that the process began in state  $m$  at time 0. This joint probability formulation is more mathematically convenient than the more natural conditional probability, as we demonstrate in the proofs of the Theorems. However, it is easy to transform  $v_{mn}(x, t)$  into the conditional probability via Bayes' rule, as we show below. Appendix A gives a preliminary lemma deriving a representation for Yule reward processes that will be useful in proving the Theorems that follow.

## 3 The distribution of phylogenetic diversity in a Yule process

The Yule process is a simple and analytically tractable mechanistic model for producing the birth times of a clade. If we assume that the species that undergoes speciation is chosen randomly

from the extant species at that time, then the Yule process is also a distribution over bifurcating trees of age  $t$ . The Markov rewards framework provides a technique to understand integrals over Yule processes in precisely this context. In this section, we study the distribution of PD in trees generated by a Yule process. PD depends only on the branching times, and not the underlying topology, of the phylogenetic tree, making it a suitable first step in our goal of integrating over trees in comparative studies.

To proceed, let  $Y(s)$  be a Yule process with branching rate  $\lambda$  that keeps track of the number of lineages at time  $s$ . We seek an expression for PD, the total branch length of the tree, which is equivalent to the area under the trajectory of the counting process  $Y(s)$ . Define a Markov reward process with  $Y(s)$  and  $a_k = k$  for  $k = 1, 2, \dots$ . Then

$$R_t = \int_0^t a_{Y(s)} ds = \int_0^t Y(s) ds. \quad (6)$$

We now state our first Theorem giving an expression for the distribution of  $R_t$  in a Yule process.

**Theorem 1.** *For a Yule process with birth rate  $\lambda$ , starting at  $Y(0) = m$  and ending at  $Y(t) = n$ ,*

$$v_{mn}(x, t) = \begin{cases} \delta(x - mt)e^{-m\lambda t} & m = n \\ \frac{\lambda^{n-m}e^{-\lambda x}}{(n-m-1)!} \sum_{j=m}^n \binom{n-1}{j-1} \binom{j-1}{m-1} (-1)^{j-m} (x-jt)^{n-m-1} H(x-jt) & n > m \end{cases} \quad (7)$$

where  $\delta(x)$  is the Dirac delta function and  $H(x)$  is the Heaviside step function.

The proof of this Theorem is given in Appendix B. There has been disagreement about whether the definition of PD in different contexts includes the root lineage (Faith, 1992; Faith and Baker, 2006; Crozier et al, 2006; Faith, 2006). We do not take a stance on this issue but note that if  $t$  is the stem age of an unresolved clade, then taking  $a_1 = 1$  in (3) includes the root in the distribution of accumulated PD, and  $a_1 = 0$  does not. The form of (7) will change slightly if  $m = 1$  and  $a_1 = 0$ .

The probability distribution of PD, conditional on  $Y(0) = m$  and  $Y(t) = n$ , is

$$f_Y(x \mid m, n, t, \lambda) = \frac{v_{mn}(x, t)}{P_{mn}(t)}, \quad (8)$$

where  $P_{mn}(t)$  is the Yule transition probability (2). This family of probability distributions has some interesting properties. Figure 2 shows  $f_Y(x|m, n, t, \lambda)$  for  $m = 1$  (with  $a_1 = 1$ ),  $n = 1, \dots, 8$ ,  $\lambda = 1.2$ , and  $t = 1$ . The unusual shape of the distribution for smaller  $n$  demonstrates the piecewise nature of the density, apparent in the functional form (7). Interestingly, Faller et al (2008) show that the distribution PD tends toward a normal distribution as  $n \rightarrow \infty$ , a fact suggested by the shape of the distributions in Figure 2.

[Figure 2 about here.]

These distributions have some practical uses. First, one can predict the PD that will arise under the Yule model from a collection of extant species up to time  $t$  in the future. Second, we can calculate the probability that future PD at time  $t$  in one group is greater than in the other, conditional on the number of species and diversification rates in both groups; this probability may have uses in conservation applications. Third, conditional on an inferred phylogenetic tree for a set of  $n$  taxa, one could compute the resulting PD  $x$  and perform a hypothesis test to evaluate the Yule-PD model using the quantity

$$\Pr(\text{PD} > x) = \int_x^{nt} f_Y(x) \, dx \quad (9)$$

where  $f_Y(x)$  is given by (8) and  $nt$  is the maximum PD that can accumulate in time  $t$ , conditional on  $Y(t) = n$ .

## 4 The distribution of expected phenotypic variance

Since researchers generally do not know the phylogenetic tree for a set of species with certainty, PD is not observable until after a tree has been estimated. Unfortunately, the uncertainty involved in estimating a tree propagates into subsequent estimates of PD based on that tree, and our distributional results may no longer apply. We therefore seek a distribution for an analogous quantity that is observable directly from knowledge of the number of species  $n$  and their trait values, bypassing the need to infer a detailed phylogenetic tree. For this, we will need a model for phenotypic trait evolution on the branches of an unknown phylogenetic tree generated by a Yule process.



The simplest and most popular model for evolution of continuous phenotypic traits on phylogenetic trees is Brownian motion (Felsenstein, 1985). Under this model, trait increments over a branch of length  $t$  are normally distributed with mean 0 and variance  $\sigma^2 t$ . The trait values for extant species at the present time are observed as the vector  $\mathbf{X} = (X_1, \dots, X_n)$ . For a given topology  $\tau$  with  $n$  taxa and branching times  $\mathbf{t} = (t_2, \dots, t_{n-1})$ , with tip data  $\mathbf{X}$  generated on the branches of this tree by zero-mean Brownian motion with variance  $\sigma^2$ , the tip data are distributed according to a multivariate normal random variable. More formally,

$$\mathbf{X} \sim N(\mathbf{0}, \sigma^2 \mathbf{C}(\tau, \mathbf{t})), \quad (10)$$

where the entries of the variance-covariance matrix  $\mathbf{C}(\tau, \mathbf{t}) = \{c_{ij}\}$  are defined as follows:  $c_{ii} = t$ , and  $c_{ij}$  is the time of shared ancestry for taxa  $i$  and  $j$ , where  $i \neq j$ . O'Meara et al (2006) introduces *disparity*, the sample variance of the tip data  $\mathbf{X}$ ,

$$\text{disparity}(\mathbf{X}) = \frac{1}{n} (\mathbf{X} - \bar{X})' (\mathbf{X} - \bar{X}), \quad (11)$$

where  $\bar{X}$  is the mean of the elements of  $\mathbf{X}$ . The expectation of the disparity, conditional on the tree topology  $\tau$ , branching times  $\mathbf{t}$ , and the Brownian variance  $\sigma^2$ , is

$$\begin{aligned} \mathbb{E}_{\mathbf{X}}(\text{disparity} \mid \tau, \mathbf{t}, \sigma^2) &= \frac{1}{n} \mathbb{E}_{\mathbf{X}}((\mathbf{X} - \bar{X})' (\mathbf{X} - \bar{X}) \mid \tau, \mathbf{t}, \sigma^2) \\ &= \sigma^2 \left[ \frac{\text{tr}(\mathbf{C}(\tau, \mathbf{t}))}{n} - \frac{1}{n^2} \mathbf{1}' \mathbf{C}(\tau, \mathbf{t}) \mathbf{1} \right] \\ &= \sigma^2 \left[ t - \frac{1}{n^2} \mathbf{1}' \mathbf{C}(\tau, \mathbf{t}) \mathbf{1} \right] \\ &= \sigma^2 \left[ t - \frac{1}{n^2} \left( nt + 2 \sum_{i=1}^n \sum_{j < i} c_{ij} \right) \right] \\ &= \sigma^2 \left[ \left( 1 - \frac{1}{n} \right) t - \frac{2}{n^2} \sum_{i=1}^n \sum_{j < i} c_{ij} \right], \end{aligned} \quad (12)$$

where we use the notation  $\mathbb{E}_{\mathbf{X}}$  to indicate that the expectation is taken over realizations of the Brownian process that generates  $\mathbf{X}$ . The fourth line above arises since the matrix  $\mathbf{C}(\tau, \mathbf{t})$  is symmetric and every element on the diagonal is  $t$ . However, every entry  $c_{ij}$  is either zero or a branching

time from the vector  $\mathbf{t}$ , so the nonzero terms in the sum consist of branching times  $t_k$ . Let  $z_k$  be the coefficient multiplying the branching time  $t_k$  in the last line of (12). Then we can express expected disparity as a weighted sum of the branching times,

$$\mathbb{E}_{\mathbf{X}}(\text{disparity} \mid \tau, \mathbf{t}, \sigma^2) = \sigma^2 \sum_{k=2}^n z_k t_k. \quad (13)$$

Figure 3 illustrates how tree topology determines the matrix  $\mathbf{C}(\tau, \mathbf{t})$  and expected disparity.

[Figure 3 about here.]

#### 4.1 Expected disparity as an accumulated reward

The expected disparity (13) has features in common with PD, since it is a scalar quantity that accumulates over the branches of the tree from time 0 to  $t$ . The difference is that disparity implicitly incorporates tree-topological factors, which enter (13) as weights in the sum of the branch lengths. In addition, a Yule tree accumulates PD even when there is a single lineage, but the same is not true for disparity. We develop these ideas in greater detail in this section.

Our goal is to express (13) as a Markov reward process in a form equivalent to (4),

$$R_t = \sigma^2 \sum_{k=1}^n a_k (t_k - t_{k-1}). \quad (14)$$

From (12) we see that  $a_n = z_n = (1 - \frac{1}{n})$ , and  $a_{n-1}$  can be found using  $a_n$  and  $z_{n-1}$ , and so on. We can formalize this recursive solution for the rewards by equating (13) and (14) as follows:

$$\sum_{k=2}^n z_k t_k = \sum_{k=1}^n a_k (t_k - t_{k-1}) = a_n t_n + \sum_{k=1}^{n-1} (a_k - a_{k+1}) t_k. \quad (15)$$

Then recursively solving for the  $a_k$ 's gives  $a_1 = 0$  and

$$a_k = \sum_{j=k}^n z_j. \quad (16)$$

for  $k = 2, \dots, n$ . Now defining  $R_t(\mathbf{a})$  to be the Yule reward process with rewards  $\mathbf{a} = (a_1, \dots, a_n)$

under the topology  $\tau$ , the expected disparity is distributed as

$$\mathbb{E}_{\mathbf{X}}(\text{disparity} \mid \tau, \lambda, \sigma^2, n) \sim \sigma^2 R_t(\mathbf{a}). \quad (17)$$

Note that we no longer need to condition on the branch lengths  $\mathbf{t} = (t_1, \dots, t_n)$  in the expected disparity – they have been “integrated out”. Therefore, to find the distribution of expected trait variance under a Brownian motion process on a Yule tree with topology  $\tau$ , we need only find the relevant rewards  $\mathbf{a}$  and compute the corresponding distribution of  $R_t(\mathbf{a})$ .

As a concrete example, consider the five-taxon tree in Figure 3. The expected disparity, given this topology  $\tau$  and arbitrary branch lengths  $\mathbf{t} = (t_2, t_3, t_4)$ , is

$$\mathbb{E}(\text{disparity} \mid \tau, \mathbf{t}, \sigma^2) = \sigma^2 \left[ \frac{4}{5}t - \frac{4}{25}t_2 - \frac{2}{25}t_3 - \frac{2}{25}t_4 \right]. \quad (18)$$

The coefficients are given by

$$\mathbf{z} = \left( -\frac{4}{25}, -\frac{2}{25}, -\frac{2}{25}, \frac{4}{5} \right). \quad (19)$$

Solving for the rewards  $\mathbf{a}$ , we obtain

$$\mathbf{a} = \left( 0, \frac{12}{25}, \frac{16}{25}, \frac{18}{25}, \frac{4}{5} \right) \quad (20)$$

which is easily verified by hand. This leads us to our second Theorem, which gives an expression for the distribution of  $R_t(\mathbf{a})$ .

**Theorem 2.** *In a Yule process with rate  $\lambda$  and arbitrary rewards  $\mathbf{a} = (a_1, \dots, a_n)$ , the Laplace transform of  $v_{mn}(x, t \mid \mathbf{a})$  is given by*

$$f_{mn}(r, t) = \begin{cases} e^{-(m\lambda + a_m r)t} & m = n, \text{ and} \\ \lambda^{n-m} \frac{(n-1)!}{(m-1)!} \sum_{j=m}^n \frac{e^{-(j\lambda + a_j r)t}}{\prod_{k \neq j} (\lambda(k-j) + r(a_k - a_j))} & n > m. \end{cases} \quad (21)$$

The proof of this Theorem is given in Appendix C. To obtain the probability distribution of the

accumulated reward, we must invert (21),

$$v_{mn}(x, t) = \mathcal{L}^{-1}[f_{mn}(r, t)](x). \quad (22)$$

For  $m = n$  and  $n = m + 1$ , there are simple expressions for the inverse Laplace transform. Under certain conditions on the rewards  $\mathbf{a}$ , there is a straightforward analytic inversion of (21) for general  $n > m$ , but the rewards computed using the times of shared ancestry in a phylogenetic tree do not always satisfy these conditions. Therefore, it is often easier to numerically invert (21); we discuss this issue in much greater detail in Appendix D, and provide a straightforward method for numerical inversion of the Laplace transform (21) based on the method popularized by Abate and Whitt (1995).

## 4.2 Approximate likelihood and inference for $\sigma^2$

We now describe a statistical procedure for using Theorem 2 to perform statistical for the unknown Brownian variance  $\sigma^2$ . Suppose that in a clade of  $n$  species we have a crude tree topology  $\tau$  (without branch lengths). This topology could be derived from parsimony, distance-based tree reconstruction methods, or one could simply use family/genus/species information to assign a hierarchy of relationships and resolve polytomies randomly. Given the tree topology  $\tau$ , one can compute the rewards  $\mathbf{a}$ . Suppose also that we have calculated trait disparity for each of  $J$  independent continuous quantitative traits that arise from Brownian motion on the branches of the unknown phylogenetic tree, starting at the root. Let

$$D_n^{(j)} = \frac{1}{n} \left( \mathbf{X}^{(j)} - \bar{X}^{(j)} \right)' \left( \mathbf{X}^{(j)} - \bar{X}^{(j)} \right), \quad (23)$$

be the observed disparity for the  $j$ th phenotypic trait, where  $\mathbf{X}^{(j)}$  is the vector of  $n$  trait values for the  $j$ th phenotypic trait and  $\bar{X}^{(j)}$  is the mean of the elements of  $\mathbf{X}^{(j)}$ . Then we calculate the mean disparity  $\bar{D}_n$  across these  $J$  traits:

$$\bar{D}_n = \frac{1}{J} \sum_{j=1}^J D_n^{(j)}. \quad (24)$$

Note that in order to find  $\bar{D}_n$ , we do not need the individual trait measurements themselves – only the disparities. Then by the law of large numbers,  $\bar{D}_n \rightarrow \mathbb{E}(D_n)$  as  $J \rightarrow \infty$ , where  $D_n$  is the

asymptotic mean disparity across all possible traits. Therefore, we approximate the distribution of  $\bar{D}_n$  as follows:

$$\begin{aligned}\bar{D}_n &\approx \mathbb{E}(D_n) \\ &\sim \mathbb{E}_{\mathbf{X}}(\text{disparity} \mid \tau, \sigma^2, \lambda, n, t) \\ &= \sigma^2 R_t(\mathbf{a}).\end{aligned}\tag{25}$$

where  $\mathbf{a}$  is the vector of rewards obtained from the topology  $\tau$ . This approximate relation provides the connection between observable mean trait disparity and the probability distribution in Theorem 2 that we need in order to compute the probability of the observed disparities. Suppose the stem age of an unresolved tree is  $t$ , and let

$$f_Y(x) = \frac{v_{mn}(x|t, \lambda, \mathbf{a})}{P_{mn}(t)}\tag{26}$$

be the distribution of expected disparity in a Yule process with general rewards  $\mathbf{a}$ , conditional on  $Y(0) = m$  and  $Y(t) = n$ . Here,  $x$  is the expected trait disparity, which we approximate by our observed (and therefore fixed)  $\bar{D}_n$ . From (25), we write

$$\frac{\bar{D}_n}{\sigma^2} \sim R_t(\mathbf{a})\tag{27}$$

so the likelihood is approximately

$$f_Y(\bar{D}_n/\sigma^2).\tag{28}$$

Finally, we propose the approximate maximum likelihood estimator

$$\hat{\sigma}^2 = \operatorname{argmax}_{\sigma^2} f_Y(\bar{D}_n/\sigma^2).\tag{29}$$

To find  $\hat{\sigma}^2$ , note that in a Yule reward process in which  $a_k < a_j$  for  $k < j$ , the value of the reward is constrained to lie in the interval

$$a_m t \leq R_t(\mathbf{a}) \leq a_n t\tag{30}$$

where we have assumed  $Y(0) = m$  and  $Y(t) = n$ . Additionally, when none of the rewards  $a_k$  are zero we are justified in dividing  $\bar{D}_n$  by the terms in (30) to obtain bounds on the possible value of  $\hat{\sigma}^2$  that maximizes (28),

$$\frac{\bar{D}_n}{a_n t} \leq \hat{\sigma}^2 \leq \frac{\bar{D}_n}{a_m t}. \quad (31)$$

When  $m = 1$  and  $a_1 = 0$ , the upper bound is infinity. However, in practice when  $m = 1$  and  $n \geq 4$ , it is safe to assume that

$$\frac{\bar{D}_n}{a_n t} \leq \hat{\sigma}^2 \leq \frac{\bar{D}_n}{a_2 t}. \quad (32)$$

We solve (29) using the numerical Newton-Raphson method provided by the R function `nlm`.

We emphasize that (29) is not a traditional maximum likelihood estimator. We have approximated the distribution of  $\bar{D}_n$  by the distribution of expected disparity, giving an approximate likelihood (28) that may not attain its maximum at the same value of  $\sigma^2$  as the true likelihood. In addition, the density (28) is non-differentiable at several points, and for  $n = m + 1$  attains its maximum at the lower boundary  $R_t(\mathbf{a}) = a_m t$  (see Figure 2 with  $m = 1$ ,  $n = 2$ ). These issues complicate application of traditional asymptotic theory for maximum likelihood estimates, and the classical large sample theory may not apply because the likelihood is only an approximation. One consequence of the violation of these traditional assumptions is that we are unable to provide meaningful standard errors for  $\hat{\sigma}^2$  using only the approximate likelihood (28).

### 4.3 Simulations

To empirically evaluate the correctness of the analytic distributions we derived in Theorem 2, we simulated trait data via Brownian motion on trees generated by a Yule process with age  $t = 1$  and branching rate  $\lambda = 1$ . For  $n = 3, \dots, 8$ , we chose one tree topology and simulated  $N_{\text{btimes}} = 2000$  sets of branch lengths from a Yule process for  $n$  species (Stadler, 2011). For each set of branching times, we simulated  $N_{\text{BM}} = 2000$  realizations of Brownian motion with  $\sigma^2 = 1$  to generate trait values at the tips of the tree (Paradis et al, 2004). For each of the 2000 sets of tip values, we calculated the mean disparity using (12). Figure 4 shows histograms of the mean disparities with the analytic distribution  $f_Y(x)$  overlaid, with good correspondence. The tree topology (with arbitrary branch lengths for display) is shown in gray above each histogram.

[Figure 4 about here.]

To evaluate our estimation methodology, we take a similar approach, but for each simulated set of mean disparities, we infer  $\hat{\sigma}^2$ , an approximate maximum likelihood estimate of  $\sigma^2$ . Figure 5 shows estimates of  $\sigma^2$  for different species richness  $n$  under different simulation conditions. For  $n = 3, \dots, 12$ , we generated 100 trees, each with  $N_{\text{btimes}} = 1, 5, 10, 100$  sets of branching times. For each set of branching times, we evolved  $N_{\text{BM}} = 1, 5, 10, 100$  traits by Brownian motion along the branches with rate  $\sigma^2 = 1$  and computed the mean disparity. Then, given the  $N_{\text{btimes}}$  mean disparities, we maximized the approximate likelihood to find  $\hat{\sigma}^2$ . Each dot in Figure 5 represents one estimate, and the dots are jittered slightly to show their density. The variance in the estimator is large when the number of simulated branching time sets and Brownian realizations is small since (25) and hence (27) become poor approximations to the mean disparity. However, the approximate maximum likelihood estimator for  $\sigma^2$  appears to have the desirable property of statistical consistency: the deviation of the estimates from the true value  $\sigma^2 = 1$  goes to zero as the number of mean disparity observations becomes large.

[Figure 5 about here.]

## 5 Application to evolution of body size in the order Carnivora

To illustrate the usefulness of our method in practical comparative inference, we estimate thirteen family-wise Brownian variance rates for body size evolution in the mammalian order Carnivora using observed log-body size disparities and species richness information (Gittleman, 1986; Gittleman and Purvis, 1998; Slater et al, 2012). Carnivora includes members with very large and small body masses, including wide diversity within individual families (Nowak and Paradiso, 1999). We included only families with 2 or more species, since families with only one species do not reveal useful information about intra-family Brownian variance. The dataset, comprising 284 species, included the families Canidae, Eupleridae, Felidae, Herpestidae, Hyaenidae, Mephitidae, Mustelidae, Otariidae, Phocidae, Prionodontidae, Procyonidae, Ursidae, and Viverridae. Figure 6 shows the backbone phylogeny (from Eizirik et al (2010)) and the unresolved clades. Our analysis takes advantage of utilities for manipulating trees and quantitative trait data in the **ape** package (described in

Paradis et al (2004)) and simulating trees with the **TreeSim** package (described in Stadler (2011)), using the statistical programming language **R** (CRAN, 2012). We intentionally limit our analysis to the Brownian variance in each family and use only body size disparity in order to demonstrate the simplicity of our approximate method under conditions of very little data.

[Figure 6 about here.]

The first step is to estimate the speciation rate  $\lambda$  from the backbone tree and the unresolved clades. Even though the tree is unobserved within each family, we can still find the exact maximum likelihood estimate of  $\lambda$  for the tree as a whole. On a fully resolved branch of length  $t$ , the log-likelihood of  $\lambda$  is  $\log(\lambda) - \lambda t$ . For an unresolved clade of age  $t$  with one species at the crown which grows to include  $n$  species, the log-likelihood from (2) is proportional to  $-\lambda t + (n - 1) \log(1 - e^{-\lambda t})$ . Summing these partial log-likelihoods over the whole tree gives the log-likelihood for  $\lambda$ ; maximizing this function, we find that  $\hat{\lambda} = 0.069$  per million years. In what follows, we assume that  $\lambda = \hat{\lambda}$ . To each unresolved family clade we associate the clade disparity for body size. Our analysis will consist of estimating the family-wise Brownian variance  $\sigma_j^2$  for the  $j$ th family, assuming one species at the stem age for each clade shown in Figure 6. In this way, we integrate over crown ages for each family under the Yule model. To apply the methodology for modeling expected trait disparity developed in section 4, we regard the unresolved clades as “soft polytomies” in which branch lengths are unknown (Purvis and Garland, 1993). We therefore resolve these polytomies randomly without assigning branch lengths.

Our analysis of the Carnivora family-wise Brownian variances takes approximately 30 seconds to run on a laptop computer. However, to evaluate the variability in our estimates, we ran the analysis 100 times with randomly resolved polytomies for each family. Table 1 shows the family name, species richness ( $n$ ), stem age ( $t$ ), observed body size disparity ( $\bar{D}_n$ ), the mean estimate of  $\sigma_j^2$ , and the approximate standard error of the estimate for each family. We calculated standard errors as the empirical standard deviation of the 100 Brownian variance estimates. Our estimates of  $\sigma^2$  reveal readily interpretable information about the evolution of body size in each family that is not available directly from observed disparities alone. For example, the families Herpestidae and Viverridae have almost identical species richness, but quite different disparity measurements. Perhaps surprisingly, we have estimated nearly equal Brownian variances for the two families. Why



does our method produce such similar estimates of  $\sigma^2$ ? The answer lies in the ages of the clades – Viverridae is almost 50% older than Herpestidae. Two clades with the same richness whose traits evolve by Brownian motion at the same rate can exhibit very different disparity measurements, depending on their ages. This example illustrates how our method provides an approximate way to untangle the complex interaction of time, species, and observed trait variance (in the words of Ricklefs (2006)) for unresolved clades.

[Table 1 about here.]

## 6 Discussion: Comparative phylogenetics without trees?

In this paper we have outlined a method for integrating over Yule trees. We presented an expression for the distribution of PD in an unresolved tree, conditional on the number of species  $n$  and age  $t$ . We showed that the expected disparity can be represented in a similar way as PD, since it accumulates along the branches of a phylogenetic tree. We also derived a statistical framework that uses a very small amount of information ( $n$ ,  $t$ , and  $\bar{D}_n$ ) for an unresolved clade to derive a meaningful estimator for the Brownian rate  $\sigma^2$ . It may seem counterintuitive that one can estimate the Brownian rate for an unresolved tree with  $n$  taxa, given a single disparity measurement. However, the structure provided by the Yule process allows this inference by providing just enough information about the distribution of branching times that generate the tree to model the average phenotypic disparity under Brownian motion. This permits analytic integration over two random objects: the collection of branching times of the tree and realizations of Brownian motion. Three assumptions make this possible: first, we fix a topology  $\tau$  without branch lengths; second, we assume that branch lengths come from a Yule process; third, we compute the distribution of expected disparity, which is a scalar quantity that encapsulates the most important information in the covariance matrix  $\mathbf{C}(\tau)$ . In exchange for these assumptions, we gain what frustrated reviewers of Felsenstein’s paper apparently wished for: an estimator that “obviate[s] the need to have an accurate knowledge of phylogeny” (Felsenstein, 1985). Whether these assumptions are warranted depends fundamentally on the scientific questions at hand, and the available data.

Perhaps the most satisfying use of our method is in providing an approximate and model-based answer to the questions posed by Ricklefs (2006) and Bokma (2010), in their similarly-named pa-

pers. Our answer is approximate because it substitutes an observation for an expectation in (25); it is model-based because we assume that trees arise from a Yule process and traits evolve Brownian motion. Equation (25) expresses a heuristic relationship explaining the origins of phenotypic disparity, which we reproduce here for emphasis:

$$\bar{D}_n \sim \sigma^2 R_t(\mathbf{a}). \quad (33)$$

On the left-hand side is the observed disparity. On the right-hand side,  $R_t(\mathbf{a})$  serves as a scalar summary of tree shape – it depends only on  $n$ , clade age  $t$ , and tree topology  $\tau$ . This reveals that even when we restrict our attention to expected disparity under the simplest evolutionary models, the interaction between  $n$ ,  $t$  and the branching structure of the tree in  $R_t(\mathbf{a})$  is complex, but the Brownian variance simply scales the tree-topological term. We see that  $\bar{D}_n$  scales linearly with  $\sigma^2$  when  $n$ ,  $t$ , and the topology  $\tau$  are fixed. However, changing one of  $n$ ,  $t$ , or  $\tau$  while holding  $\sigma^2$  constant will induce a nonlinear change in  $\bar{D}_n$ . We conclude that it is not possible to partition the time-dependent and speciation-dependent influences on the accumulation of trait variance in a simple way as suggested by Ricklefs (2006) under the stochastic models we study in this paper.

As an inducement to spur research on analytic integration over trees, Bokma (2010) offers a monetary reward for an expression for the distribution of sample variance from a birth-death tree with Brownian trait evolution on its branches. We have solved a simpler version of Bokma’s challenge by providing the distribution of expected trait variance for a specific topology under a pure-birth process. The expression Bokma (2010) seeks is difficult to find for two reasons: first, it would require analytic integration over discrete tree topologies; second, and more intuitively, integrating over both topologies and Brownian realizations would subsume the Brownian variance  $\sigma^2$  on the right-hand side of (25) into a nonlinear term that depended on  $n$ ,  $t$ , and  $\sigma^2$  in a very complicated way. Alternatively, simulation-based approaches provide an appealing alternative method to integrate over trees and Brownian motions without requiring approximation of the disparity by its expectation. Indeed, Bayesian methods exist to sample from the distribution of Yule trees, conditional on observed trait values at the tips, thereby providing both estimates of Brownian rates and phylogenies simultaneously, while using all the available trait data (Drummond et al, 2012).

Tree-free comparative evolutionary biology comes at a price – there are several important draw-

backs to our approach. First, even under the Yule model for speciation (with the correctly specified branching rate  $\lambda$ ) and zero-mean Brownian motion for traits, integrating over all possible Yule trees introduces great uncertainty in estimates of  $\sigma^2$ . Figure 5 illustrates this issue: while the estimates of  $\sigma^2$  eventually converge to the true value as the number of branch length and trait realizations becomes large, the variance in these estimates can be substantial for smaller datasets. Furthermore, the assumption that  $\bar{D}_n \sim \sigma^2 R_t(\mathbf{a})$  may be suspect if the number of traits analyzed is small enough that the mean trait disparity is a poor substitute for the expected disparity.

We conclude with a mixed message about analytic integration over trees. First, it is possible to derive meaningful estimators for parameters of interest under simple evolutionary models, if one is willing to make assumptions about the mean behavior of the models. The estimates are usually reasonable, and may provide valuable insight into the basic properties of evolutionary change under these models – even our simplistic analysis of Carnivora body size evolution reveals the complex interaction of clade age, species number, and evolutionary rate. These estimates may be useful as starting points for more time-consuming simulation analyses. Second, and more pessimistically, sophisticated analytic methods for integrating over trees cannot conjure evolutionary information from the data that is not there already. As evolutionary biologists further refine our knowledge of the tree of life, the number of clades whose phylogeny is truly unknown may diminish, along with interest in tree-free estimation methods.

## Acknowledgements

We are extremely grateful to Michael Alfaro and Graham Slater for introducing us to the problem of finding the distribution of quantitative trait variance, providing the Carnivora dataset, and for helpful comments on the manuscript. John Welch and Eric Stone provided insightful criticism and suggestions. FWC was supported by NIH grant T32GM008185, and MAS was supported by NIH grants R01 GM086887, HG006139, and NSF grant DMS-0856099.

## A Markov rewards for Yule processes

In this Appendix, we prove one lemma and the two Theorems presented in the text. In the first proof, we derive a representation of the forward equation for a Yule reward process. Our

development follows that given by Neuts (1995).

**Lemma 1.** *In a Yule reward process  $R_t = \int_0^t a_{Y(s)} \, ds$  with arbitrary positive rewards  $a_1, a_2, \dots$ , the Laplace-transformed reward probabilities satisfy the ordinary differential equations*

$$\frac{df_{mn}(r, t)}{dt} = -(n\lambda + a_n r)f_{mn}(r, t) + (n-1)\lambda f_{m, n-1}(r, t). \quad (34)$$

*Proof.* Let  $V_{mn}(x, t) = \Pr(R_t \leq x, Y(t) = n \mid Y(0) = m)$ . We can re-write this quantity in a more useful form by conditioning on the time of departure  $u$  from state  $m$ , noting that the accumulated reward is  $a_m u$ , and then integrating over  $u$ . If  $m = n$  and no departure occurs, the accumulated reward is  $a_m t$ . Putting these ideas together, we obtain

$$\begin{aligned} V_{mn}(x, t) &= \Pr(R_t \leq x, Y(s) = m \text{ for } 0 \leq s \leq t) \\ &\quad + \int_0^t \Pr(m \rightarrow m+1 \text{ at time } u) V_{m+1, n}(x - a_m u, t - u) \, du \\ &= \delta_{mn} e^{-m\lambda t} H(x - a_m t) + \int_0^t m\lambda e^{-m\lambda u} V_{m+1, n}(x - a_m u, t - u) \, du. \end{aligned} \quad (35)$$

Now consider the Laplace transform  $F_{mn}(r, t)$  of  $V_{mn}(x, t)$ , with respect to the reward variable  $x$ ,

$$\begin{aligned} F_{mn}(r, t) &= \mathcal{L}[V_{mn}(x, t)](r) \\ &= \int_0^\infty e^{-rx} V_{mn}(x, t) \, dx \\ &= \delta_{mn} \frac{e^{-(m\lambda + a_m r)t}}{r} + \int_0^t m\lambda e^{-m\lambda u} \int_{a_m u}^\infty e^{-rx} V_{m+1, n}(x - a_m u, t - u) \, dx \, du. \end{aligned} \quad (36)$$

Making the substitution  $y = x - a_m u$  in the Laplace integral, we have

$$\begin{aligned} F_{mn}(s, t) &= \delta_{mn} \frac{e^{-(m\lambda + a_m r)t}}{r} + \int_0^t m\lambda e^{-m\lambda u} \int_0^\infty e^{-r(y + a_m u)} V_{m+1, n}(y, t - u) \, dy \, du \\ &= \delta_{mn} \frac{e^{-(m\lambda + a_m r)t}}{r} + \int_0^t m\lambda e^{-(m\lambda + a_m r)u} F_{m+1, n}(r, t - u) \, du. \end{aligned} \quad (37)$$

Now multiplying both sides by  $e^{(m\lambda + a_m r)t}$  and differentiating with respect to  $t$ , we obtain

$$\begin{aligned} \frac{\partial}{\partial t} \left[ e^{(m\lambda + a_m r)t} F_{mn}(r, t) \right] &= \frac{\partial}{\partial t} \int_0^t m\lambda e^{(m\lambda + a_m r)(t-u)} F_{m+1, n}(r, t - u) \, du \\ &= \frac{\partial}{\partial t} \int_0^t m\lambda e^{(m\lambda + a_m r)u} F_{m+1, n}(r, u) \, du. \end{aligned} \quad (38)$$

Expanding the left-hand side by the product rule and using the fundamental theorem of calculus on the right, we find that

$$e^{(m\lambda+a_m r)t} \left( (m\lambda + a_m r)F_{mn}(r, t) + \frac{\partial}{\partial t} F_{mn}(r, t) \right) = e^{(m\lambda+a_m r)t} m\lambda F_{m+1,n}(r, t). \quad (39)$$

Cancelling common factors and rearranging, we obtain the Kolmogorov backward equation,

$$\frac{\partial}{\partial t} F_{mn}(r, t) = -(m\lambda + a_m r)F_{mn}(r, t) + m\lambda F_{m+1,n}(r, t). \quad (40)$$

However,

$$rF_{mn}(r, t) = \mathcal{L} \left[ \frac{\partial}{\partial x} V_{mn}(x, t) \right] (r) = \mathcal{L} [v_{mn}(x, t)](r) = f_{mn}(r, t). \quad (41)$$

Plugging  $rF_{mn}(r, t) = f_{mn}(r, t)$  into (40), we find that the  $f_{mn}(r, t)$  satisfy the same system of ordinary differential equations,

$$\frac{\partial}{\partial t} f_{mn}(r, t) = -(m\lambda + a_m r)f_{mn}(r, t) + m\lambda f_{m+1,n}(r, t). \quad (42)$$

These are the *backward equations* for the Laplace transformed reward process. To solve (42), we note that any solution to the forward equations is a solution to the backward equations in a birth process (Grimmett and Stirzaker, 2001). Therefore, (42) is equivalent to the forward system

$$\frac{\partial f_{mn}(r, t)}{\partial t} = -(n\lambda + a_n r)f_{mn}(r, t) + (n-1)\lambda f_{m,n-1}(r, t) \quad (43)$$

for  $n = m, m+1, m+2, \dots$ . This completes the proof.  $\square$

## B Proof of Theorem 1

*Proof.* Lemma 1 with  $a_k = k$  for  $k = 0, 1, \dots$  gives

$$\frac{\partial f_{mn}(r, t)}{\partial t} = -n(\lambda + r)f_{mn}(r, t) + (n-1)\lambda f_{m,n-1}(r, t). \quad (44)$$

Define  $g_{mn}(r, s)$  to be the Laplace transform of  $f_{mn}(r, t)$  with respect to the time variable  $t$ . Transforming (44) gives

$$sg_{mn}(r, s) - \delta_{mn} = -n(\lambda + r)g_{mn}(r, s) + (n - 1)\lambda g_{m, n-1}(r, s). \quad (45)$$

Letting  $m = n$ , we find that

$$g_{nn}(r, s) = \frac{1}{s + n(\lambda + r)}. \quad (46)$$

Next, we form a recurrence and solve for  $g_{mn}(r, s)$  to obtain

$$\begin{aligned} g_{mn}(r, s) &= \frac{(n - 1)\lambda}{s + n(\lambda + r)} g_{m, n-1}(r, s) \\ &= \frac{(n - 1) \cdots m \lambda^{n-m}}{\prod_{j=m+1}^n [s + j(\lambda + r)]} g_{m, m+1}(r, s) \\ &= \frac{(n - 1)!}{(m - 1)!} \frac{\lambda^{n-m}}{\prod_{j=m}^n [s + j(\lambda + r)]}. \end{aligned} \quad (47)$$

We proceed via a partial fractions decomposition of the product in the denominator above,

$$\begin{aligned} g_{mn}(r, s) &= \lambda^{n-m} \frac{(n - 1)!}{(m - 1)!} \sum_{j=m}^n \left( \prod_{k \neq j} (\lambda + r)(k - j) \right)^{-1} \frac{1}{s + j(\lambda + r)} \\ &= \frac{(n - 1)!}{(m - 1)!} \lambda^{n-m} \sum_{j=m}^n \frac{\left[ \left( \prod_{k=m}^{j-1} (k - j) \right) \left( \prod_{k=j+1}^n (k - j) \right) \right]^{-1}}{(\lambda + r)^{n-m}} \frac{1}{s + j(\lambda + r)} \\ &= \frac{(n - 1)!}{(m - 1)!} \lambda^{n-m} \sum_{j=m}^n \frac{\left[ (-1)^{j-m} (j - m)! (n - j)! \right]^{-1}}{(\lambda + r)^{n-m}} \frac{1}{s + j(\lambda + r)} \\ &= \lambda^{n-m} \sum_{j=m}^n \binom{n - 1}{j - 1} \binom{j - 1}{m - 1} \frac{(-1)^{j-m}}{(\lambda + r)^{n-m}} \frac{1}{s + j(\lambda + r)}. \end{aligned} \quad (48)$$

when  $n > m$ . Inverse transforming with respect to  $s$ , we obtain

$$f_{mm}(r, t) = e^{-m(\lambda + r)t} \quad (49)$$

and

$$f_{mn}(r, t) = \lambda^{n-m} \sum_{j=m}^n \binom{n - 1}{j - 1} \binom{j - 1}{m - 1} \frac{(-1)^{j-m}}{(\lambda + r)^{n-m}} e^{-j(\lambda + r)t} \quad (50)$$

when  $n > m$ . Again inverse transforming (50), this time with respect to the Laplace reward variable  $r$ , we find that for  $m = n$ ,

$$v_{mm}(x, t) = \delta(x - mt)e^{-m\lambda t} \quad (51)$$

which is a point mass at  $x = mt$ . For  $n > m$ ,

$$v_{mn}(x, t) = \frac{\lambda^{n-m}e^{-\lambda x}}{(n-m-1)!} \sum_{j=m}^n \binom{n-1}{j-1} \binom{j-1}{m-1} (-1)^{j-m} (x-jt)^{n-m-1} H(x-jt). \quad (52)$$

This completes the proof.  $\square$

## C Proof of Theorem 2

*Proof.* Lemma 1 with arbitrary rewards  $a_k$ ,  $k = 1, 2, \dots$ , gives

$$\frac{df_{mn}(r, t)}{dt} = -(n\lambda + a_n r)f_{mn}(r, t) + (n-1)\lambda f_{m, n-1}(r, t). \quad (53)$$

To solve the system, apply the Laplace transform with respect to time  $t$ . First note that the transform of  $f_{mn}(r, t)$  is

$$g_{mn}(r, s) = \frac{1}{s + m\lambda + a_m r}. \quad (54)$$

Transforming the  $n$ th equation, and recalling that  $f_{mn}(r, 0) = 0$  for  $n > m$ ,

$$\begin{aligned} sg_{mn}(r, s) - f_{mn}(r, 0) &= -(n\lambda + a_n r)g_{mn}(r, s) + (n-1)\lambda g_{m, n-1}(r, s) \\ g_{mn}(r, s)(s + n\lambda + a_n r) &= (n-1)\lambda g_{m, n-1}(r, s) \\ g_{mn}(r, s) &= \frac{(n-1)\lambda}{s + n\lambda + a_n r} g_{m, n-1}(r, s) \\ &= \frac{(n-1)!}{(m-1)!} \frac{\lambda^{n-m}}{\prod_{j=m+1}^n (s + j\lambda + a_j r)} g_{mm}(r, s) \\ &= \frac{(n-1)!}{(m-1)!} \frac{\lambda^{n-m}}{\prod_{j=m}^n (s + j\lambda + a_j r)}. \end{aligned} \quad (55)$$

We expand the denominator by partial fractions to find

$$g_{mn}(r, s) = \lambda^{n-m} \frac{(n-1)!}{(m-1)!} \sum_{j=m}^n \frac{\prod_{k \neq j} [\lambda(k-j) + r(a_k - a_j)]^{-1}}{s + j\lambda + a_j r}. \quad (56)$$

Transforming back to the time domain, we have, for  $m = n$ ,

$$f_{mm}(r, t) = e^{-(m\lambda + a_m r)t}. \quad (57)$$

When  $n > m$ ,

$$f_{mn}(r, t) = \lambda^{n-m} \frac{(n-1)!}{(m-1)!} \sum_{j=m}^n \frac{e^{-(j\lambda + a_j r)t}}{\prod_{k \neq j} (\lambda(k-j) + r(a_k - a_j))}. \quad (58)$$

This completes the proof.  $\square$

## D Analytic and numerical inversion

Analytic inversion of (21) in Theorem 2 is possible, but unfortunately depends on the structure of the tree topology in unexpected ways. One convenient property of the rewards  $\mathbf{a} = (a_1, \dots, a_n)$  is that  $a_i < a_{i+1}$  for all  $1 \leq i \leq n-1$ , a fact apparent from (12). Therefore  $a_i \neq a_j$  for distinct  $i$  and  $j$ . When  $m = n$ , no speciation events have taken place, and we have

$$v_{mm}(x, t) = \delta(x - a_m t) e^{-m\lambda t}. \quad (59)$$

For  $n = m + 1$ , there is only one distinct topology, so

$$\begin{aligned} v_{m,m+1}(x, t) = \frac{m\lambda e^{-m\lambda t}}{a_{m+1} - a_m} & \left[ \exp\left(-\frac{\lambda(x - a_m t)}{a_{m+1} - a_m}\right) H(x - a_m t) \right. \\ & \left. - e^{\lambda t} \exp\left(-\frac{\lambda(x - a_{m+1} t)}{a_{m+1} - a_m}\right) H(x - a_{m+1} t) \right]. \end{aligned} \quad (60)$$

In general, when

$$\frac{\ell - j}{a_\ell - a_j} - \frac{k - j}{a_k - a_j} \neq 0 \quad (61)$$

for any  $l, k$ , or  $j$  in  $1, \dots, n$ , then (58) becomes

$$\begin{aligned} f_{mn}(r, t) &= \lambda^{n-m} \frac{(n-1)!}{(m-1)!} \sum_{j=m}^n \frac{e^{-(j\lambda + a_j r)t}}{\prod_{k \neq j} (a_k - a_j) \left( \frac{\lambda(k-j)}{a_k - a_j} + r \right)} \\ &= \lambda^{n-m} \frac{(n-1)!}{(m-1)!} \sum_{j=m}^n \frac{e^{-(j\lambda + a_j r)t}}{\prod_{k \neq j} (a_k - a_j)} \sum_{\substack{\ell \neq k \\ \ell \neq j}} \frac{\left( \frac{\lambda(\ell-j)}{a_\ell - a_j} - \frac{\lambda(k-j)}{a_k - a_j} \right)^{-1}}{\frac{\lambda(k-j)}{a_k - a_j} + r}, \end{aligned} \quad (62)$$



and so the full probability density for  $n > m$  is

$$v_{mn}(x, t) = \lambda^2 \frac{(n-1)!}{(m-1)!} \sum_{j=m}^n \frac{e^{-j\lambda t} H(x - a_j t)}{\prod_{k \neq j} (a_k - a_j)} \sum_{k \neq j} \frac{\exp \left[ -\frac{\lambda(k-j)}{a_k - a_j} (x - a_j t) \right]}{\prod_{\substack{\ell \neq k \\ \ell \neq j}} \left( \frac{\ell-j}{a_\ell - a_j} - \frac{k-j}{a_k - a_j} \right)} \quad (63)$$

However, the rewards  $\mathbf{a}$  for many topologies do not satisfy (61). This can be seen in Figure 7, where  $m = 2$  and  $n = 4$ . Then we see that when  $j = 2$ ,  $k = 4$ , and  $\ell = 3$  in (63),

$$\frac{4-2}{a_4 - a_2} = \frac{4-2}{0.75 - 0.5} = 8 \quad (64)$$

and

$$\frac{3-2}{a_3 - a_2} = \frac{3-2}{0.75 - 0.625} = 8, \quad (65)$$

so the denominator in the second sum in (63) is zero. Unfortunately this happens whenever there is symmetry in the tree so that more than one pair of taxa have the same time of shared ancestry. Note also that (63) does not reduce to (21) in Theorem 2 when  $a_k = k$  since the denominator in the summand of (63) is zero.

[Figure 7 about here.]

Despite the difficulty in writing a general inversion to obtain  $f_{mn}(x, t)$  for any topology, numerical inversion to arbitrary precision remains straightforward. Abate and Whitt (1995) describe a numerical method for inverting the Laplace transform of probability densities by a discrete Riemann sum using the trapezoidal rule with step size  $h$ :

$$v_{mn}(x, t) \approx \frac{e^{A/2}}{2x} \operatorname{Re} \left[ f_{m,n} \left( \frac{A}{2x}, t \right) \right] + \frac{e^{A/2}}{x} \sum_{k=1}^{\infty} (-1)^k \operatorname{Re} \left[ f_{m,n} \left( \frac{A + 2k\pi i}{2x}, t \right) \right], \quad (66)$$

where we choose  $A = 20$ .

## References

Abate J, Whitt W (1995) Numerical inversion of Laplace transforms of probability distributions. ORS J Comput 7(1):36–43

- Alfaro M, Santini F, Brock C, Alamillo H, Dornburg A, Rabosky D, Carnevale G, Harmon L (2009) Nine exceptional radiations plus high turnover explain species diversity in jawed vertebrates. *PNAS* 106(32):13,410–13,414
- Bailey NTJ (1964) *The Elements of Stochastic Processes with Applications to the Natural Sciences*. Wiley New York
- Blomberg S, Garland T, Ives A (2003) Testing for phylogenetic signal in comparative data: behavioral traits are more labile. *Evolution* 57(4):717–745
- Bokma F (2010) Time, species, and separating their effects on trait variance in clades. *Syst Biol* 59(5):602–607
- CRAN (2012) The comprehensive R archive network. URL <http://cran.r-project.org>
- Crozier R, Agapow P, Dunnett L (2006) Conceptual issues in phylogeny and conservation: a reply to faith and baker. *Evol Bioinform Online* 2:197–199
- Drummond AJ, Suchard MA, Xie D, Rambaut A (2012) Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution*
- Eizirik E, Murphy W, Koepfli K, Johnson W, Dragoo J, Wayne R, O’Brien S (2010) Pattern and timing of diversification of the mammalian order carnivora inferred from multiple nuclear gene sequences. *Mol Phylogenet Evol* 56(1):49–63
- Eldredge N, Gould S (1972) Punctuated equilibria: an alternative to phyletic gradualism, Freeman, Cooper and Company, pp 82–115
- Faith D (1992) Conservation evaluation and phylogenetic diversity. *Biol Conserv* 61(1):1–10
- Faith D (2006) The role of the phylogenetic diversity measure, pd, in bio-informatics: getting the definition right. *Evolutionary bioinformatics online* 2:277
- Faith D, Baker A (2006) Phylogenetic diversity (pd) and biodiversity conservation: some bioinformatics challenges. *Evol Bioinform Online* 2:121
- Faller B, Pardi F, Steel M (2008) Distribution of phylogenetic diversity under random extinction. *J Theor Biol* 251(2):286–296

- Felsenstein J (1985) Phylogenies and the comparative method. *Am Nat* 125(1):1–15
- Foote M (1993) Contributions of individual taxa to overall morphological disparity. *Paleobiology* 19(4):403–419
- Garland T, Harvey P, Ives A (1992) Procedures for the analysis of comparative data using phylogenetically independent contrasts. *Syst Biol* 41(1):18–32
- Gernhard T, Hartmann K, Steel M (2008) Stochastic properties of generalised Yule models, with biodiversity applications. *J Math Biol* 57(5):713–735
- Gittleman J (1986) Carnivore life history patterns: allometric, phylogenetic, and ecological associations. *Am Nat* 127(6):744–771
- Gittleman J, Purvis A (1998) Body size and species–richness in carnivores and primates. *P Roy Soc Lond B Bio* 265(1391):113–119
- Gould S, Eldredge N (1977) Punctuated equilibria: the tempo and mode of evolution reconsidered. *Paleobiology* 3(2):115–151
- Grafen A (1989) The phylogenetic regression. *Phil T Roy Soc B* 326(1233):119–157
- Grimmett G, Stirzaker D (2001) Probability and random processes. Oxford University Press
- Harvey P, Pagel M (1991) The comparative method in evolutionary biology. Oxford university press
- Martins E, Hansen T (1997) Phylogenies and the comparative method: a general approach to incorporating phylogenetic information into the analysis of interspecific data. *Am Nat* 149(4):646–667
- Minin V, Suchard M (2008) Counting labeled transitions in continuous-time markov models of evolution. *J Math Biol* 56(3):391–412
- Mooers A, Gascuel O, Stadler T, Li H, Steel M (2011) Branch lengths on birth-death trees and the expected loss of phylogenetic diversity. *Syst Biol* 61(2):195–203
- Moritz C (2002) Strategies to protect biological diversity and the evolutionary processes that sustain it. *Syst Biol* 51(2):238–254

- Mulder W (2011) Probability distributions of ancestries and genealogical distances on stochastically generated rooted binary trees. *J Theor Biol* 280(1):139–145
- Nee S (2006) Birth-death models in macroevolution. *Annu Rev Ecol Evol S* 37:1–17
- Nee S, May RM, Harvey PH (1994) The reconstructed evolutionary process. *Philos T Roy Soc B* 344(1309):305–311
- Neuts MF (1995) *Algorithmic Probability: A Collection of Problems (Stochastic Modeling Series)*. Chapman and Hall/CRC
- Nowak R, Paradiso J (1999) *Walker’s mammals of the world*. Cambridge Univ Press
- O’Meara BC, Ané C, Sanderson MJ, Wainwright PC (2006) Testing for different rates of continuous trait evolution using likelihood. *Evolution* 60(5):922–933
- Paradis E, Claude J, Strimmer K (2004) APE: analyses of phylogenetics and evolution in R language. *Bioinformatics* 20(2):289–290
- Purvis A (2004) Evolution: How do characters evolve? *Nature* 432(7014):Published online
- Purvis A, Garland T (1993) Polytomies in comparative analyses of continuous characters. *Syst Biol* 42(4):569–575
- Rannala B, Yang Z (1996) Probability distribution of molecular evolutionary trees: a new method of phylogenetic inference. *J Mol Evol* 43(3):304–311
- Revell L (2010) Phylogenetic signal and linear regression on species data. *Methods in Ecology and Evolution* 1(4):319–329
- Ricklefs R (2004) Cladogenesis and morphological diversification in passerine birds. *Nature* 430(6997):338–341
- Ricklefs R (2006) Time, species, and the generation of trait variance in clades. *Syst Biol* 55(1):151–159
- Sidlauskas B (2007) Testing for unequal rates of morphological diversification in the absence of a detailed phylogeny: a case study from characiform fishes. *Evolution* 61(2):299–316

- Slater G, Harmon L, Wegmann D, Joyce P, Revell L, Alfaro M (2012) Fitting models of continuous trait evolution to incompletely sampled comparative data using Approximate Bayesian Computation. *Evolution* 66:752–762
- Stadler T (2011) Simulating trees with a fixed number of extant species. *Syst Biol* 60(5):676–684
- Stadler T, Steel M (2012) Distribution of branch lengths and phylogenetic diversity under homogeneous speciation models. *J Theor Biol* 297:33–40
- Steel M, McKenzie A (2001) Properties of phylogenetic trees generated by Yule-type speciation models. *Math Biosci* 170(1):91–112
- Steel M, McKenzie A (2002) The ‘shape’ of phylogenies under simple random speciation models. *Biological Evolution and Statistical Physics* 585:162–180
- Steel M, Mooers A (2010) The expected length of pendant and interior edges of a Yule tree. *Appl Math Lett* 23(11):1315–1319
- Stone E (2011) Why the phylogenetic regression appears robust to tree misspecification. *Syst Biol* 60(3):245–260
- Turnbaugh P, Hamady M, Yatsunenko T, Cantarel B, Duncan A, Ley R, Sogin M, Jones W, Roe B, Affourtit J, et al (2008) A core gut microbiome in obese and lean twins. *Nature* 457(7228):480–484
- Webb C, Ackerly D, McPeck M, Donoghue M (2002) Phylogenies and community ecology. *Annu Rev Ecol Syst* 33:475–505
- Yule (1925) A mathematical theory of evolution, based on the conclusions of Dr. J. C. Willis, F.R.S. *Philos T R Soc Lon B* 213:21–87

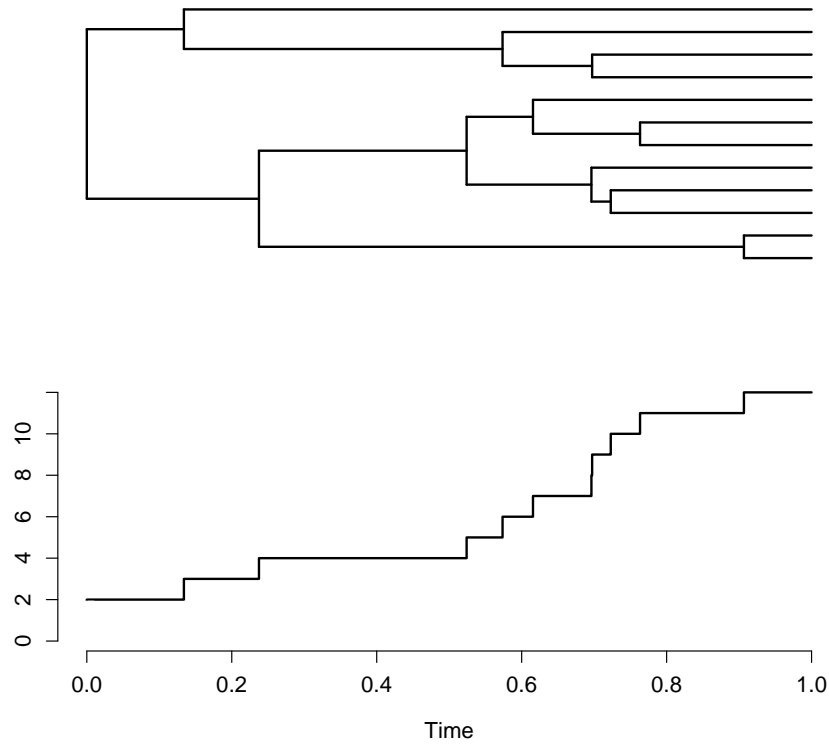


Figure 1: Example of a Yule (pure-birth) tree with the corresponding counting process  $Y(t)$  below. The birth rate in this example is  $\lambda = 2$ . In the counting process representation of this realization, we only keep track of the total number of species in existence at each time.

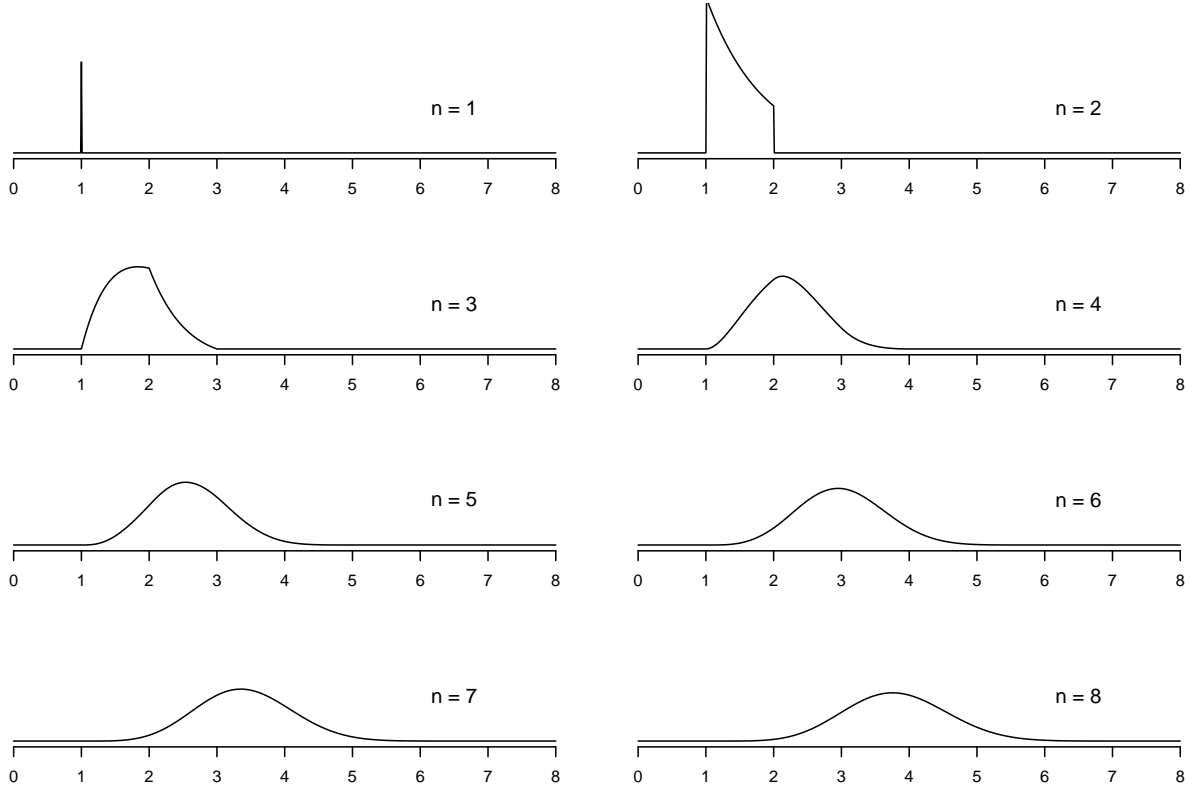


Figure 2: Probability densities of phylogenetic diversity (PD), or total branch length, under the Yule process starting at  $Y(0) = 1$ , ending at  $Y(t) = n$  for  $n = 1, \dots, 8$ , with  $t = 1$  and  $\lambda = 1.2$ . When  $n = 1$ , no births have occurred, so the accrued PD must be exactly  $x = t = 1$ , which we represent here as a point mass at 1. For  $n = 2$ , the minimum accumulated PD is one, since the process spent at most one unit of time with one species; likewise the maximum accumulated PD is two, since the process spent at most one unit of time with two species. The functional form of (7) reveals the piecewise nature of the density, which gradually becomes smoother as  $n$  becomes large. The vertical probability axis is the same for all plots.

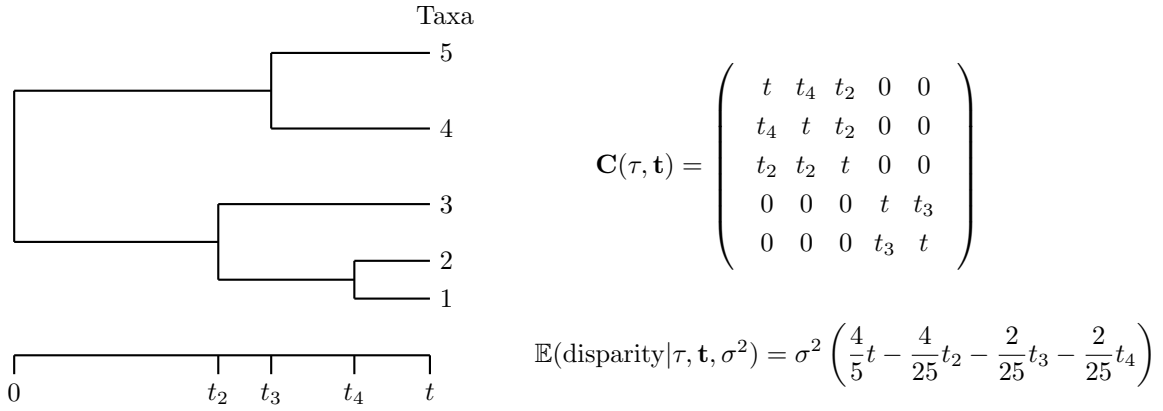


Figure 3: How tree topology determines the matrix of Brownian covariances and expected disparity. At left, a tree topology  $\tau$  of crown age  $t$  has 5 taxa and branch times  $\mathbf{t} = (t_2, t_3, t_4)$ , where  $t_k$  is the time of the branch from  $k$  to  $k + 1$  lineages. At right, the corresponding matrix  $\mathbf{C}(\tau, \mathbf{t})$  of Brownian covariances. The diagonal entries of  $\mathbf{C}(\tau, \mathbf{t})$  are all  $t$ . The  $(i, j)$ th entry of  $\mathbf{C}(\tau, \mathbf{t})$  is the time of shared ancestry between taxa  $i$  and  $j$ , for  $i \neq j$ . For example, taxa 1 and 2 share ancestry for time  $t_4$ . Expected disparity (using Brownian variance  $\sigma^2$ ) is calculated using (12). Trait disparity cannot accumulate when there is only one species, so we draw the tree  $\tau$  beginning with two lineages at time 0.



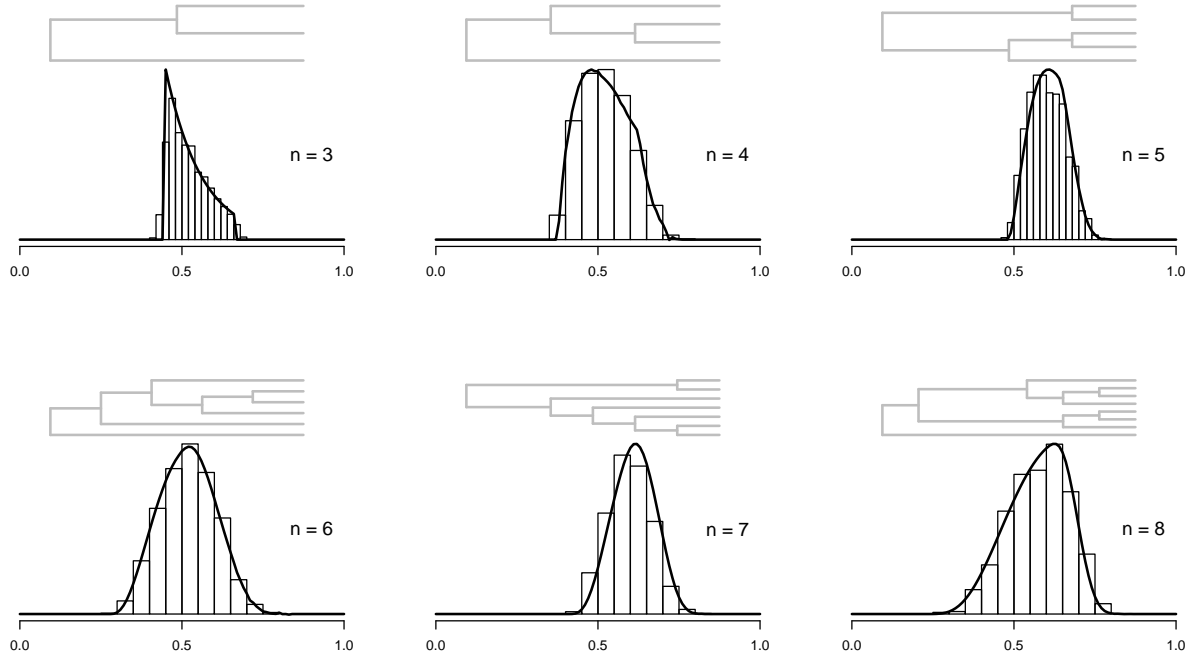


Figure 4: Empirical correspondence between the derived expressions for the distribution of expected disparity and simulated mean disparity histograms for trees with different numbers of taxa. For each  $n = 3, \dots, 8$ , we simulated a single tree topology (shown above each histogram in gray). We then simulated 2000 sets of branch lengths for this topology under the Yule process. For each set of branch lengths, we calculated the mean disparity from 2000 simulations of zero-mean Brownian motion with variance  $\sigma^2 = 1$  on this tree.

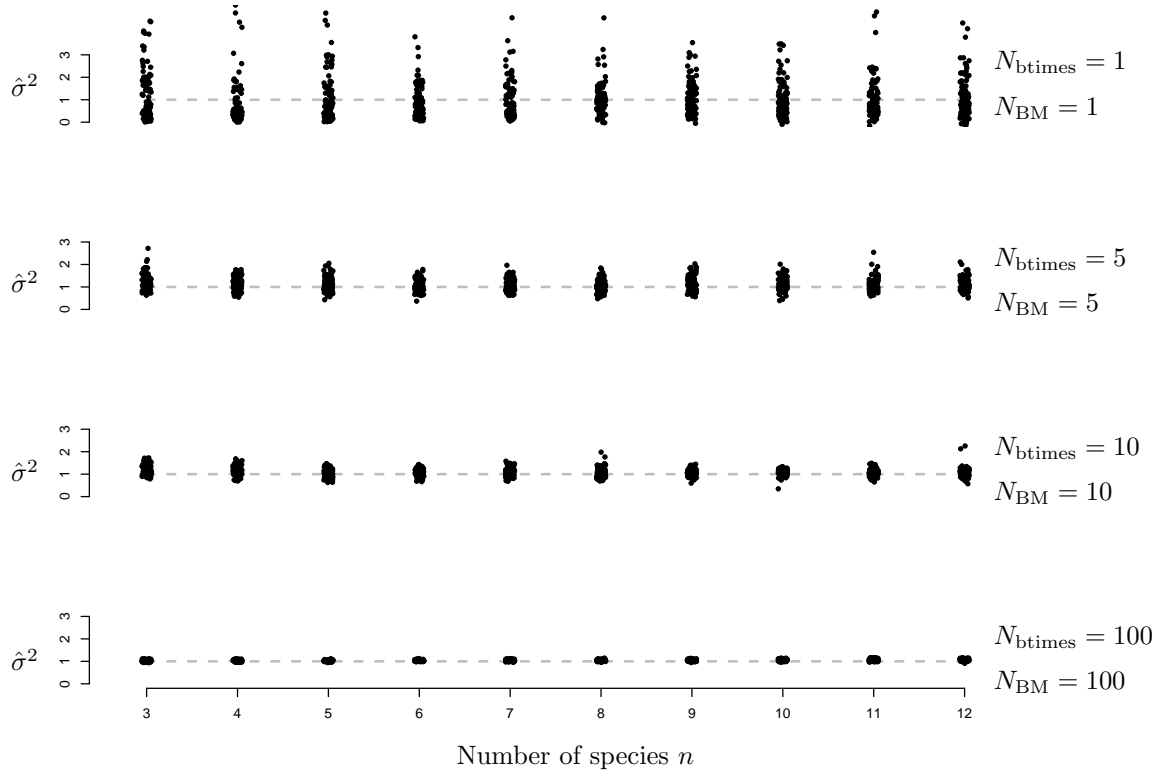


Figure 5: Empirical consistency of approximate maximum likelihood estimates of  $\sigma^2$ . The number of species  $n$  in the unobserved phylogenetic tree is shown on the horizontal axis. Each set of plots shows 100 estimates of  $\sigma^2$  from  $N_{\text{btimes}}$  simulations of different branching times under a Yule process with  $n$  species and from  $N_{\text{BM}}$  independent realizations of Brownian motion used to compute the mean disparity for each set of branching times. The estimates are jittered to show the sampling distribution. The gray dotted line shows the true value  $\sigma^2 = 1$ .

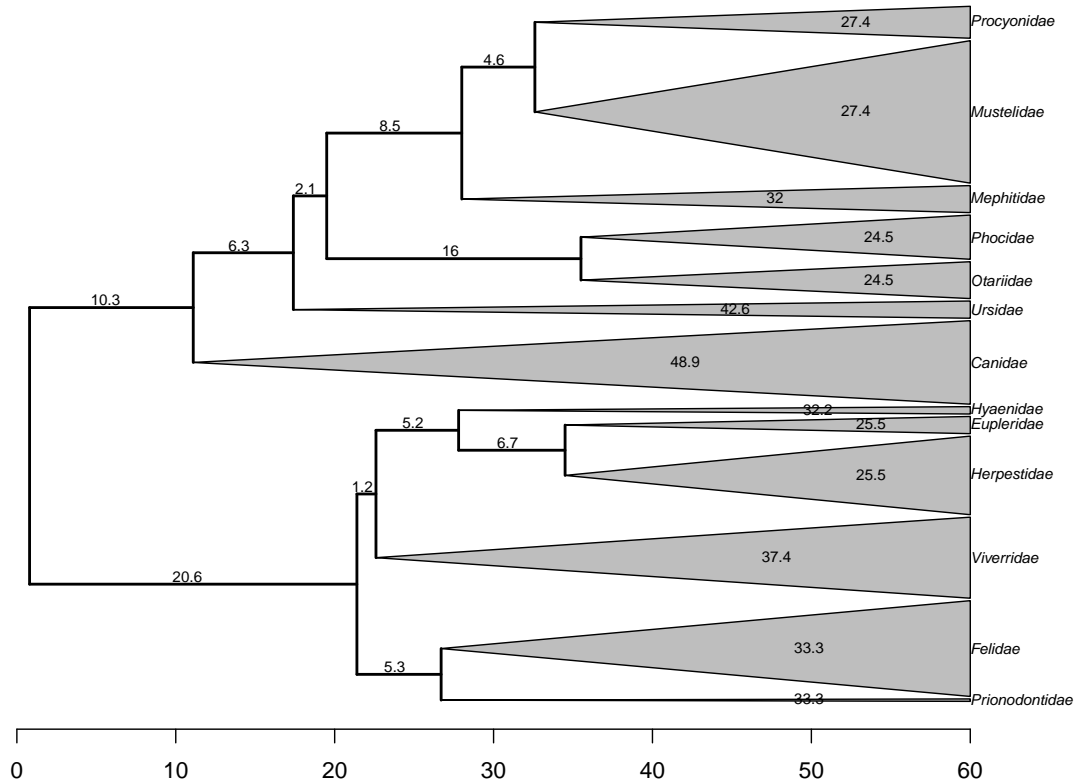


Figure 6: A family-level phylogenetic tree for order Carnivora. The phylogeny within each family, shown here as a gray triangle, is not known with certainty. The length of the base of the gray triangles represents the number of species in the family. The “backbone” tree connecting the unresolved clades is assumed fixed. Branch lengths are shown along each branch.

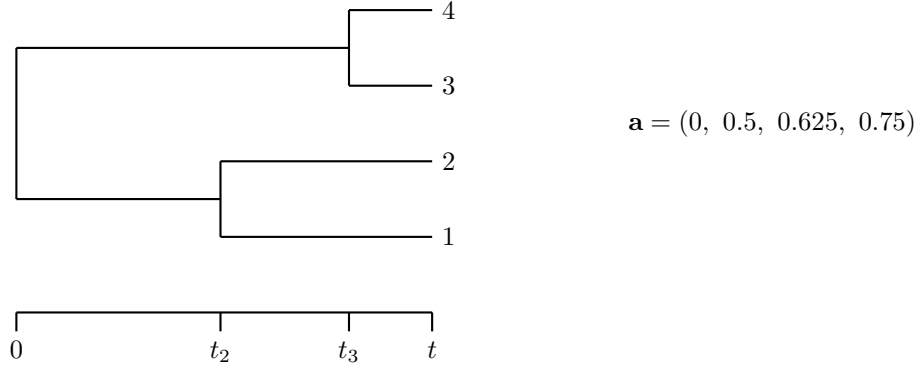


Figure 7: Demonstration of a problematic reward vector  $\mathbf{a}$  computed for the symmetric four-taxon tree. In this case, the analytic inversion formula (63) cannot be applied, since the denominator in the sum becomes zero.

Family	Richness	TMRCa	Disparity	$\hat{\sigma}^2$	SE
Prionodontidae	2	33.30	0.131	0.051	0
Felidae	40	33.30	1.588	0.118	0.094
Viverridae	34	37.40	0.606	0.034	0.018
Herpestidae	33	25.50	0.482	0.035	0.017
Eupleridae	8	25.50	0.916	0.081	0.010
Hyaenidae	4	32.20	0.805	0.084	0.001
Canidae	35	48.90	0.678	0.031	0.012
Ursidae	8	42.60	0.303	0.025	0.002
Otariidae	16	24.50	0.386	0.029	0.005
Phocidae	19	24.50	0.751	0.065	0.030
Mephitidae	12	32.00	0.570	0.038	0.004
Mustelidae	59	27.40	2.263	0.220	0.239
Procyonidae	14	27.40	0.531	0.038	0.006

Table 1: Species richness, TMRCa, body size disparity, and estimated Brownian variance  $\hat{\sigma}^2$  for each family in the Carnivora dataset. Note that Canidae and Herpestidae have very different disparity measurements, but nearly identical estimates of  $\sigma^2$ . This discrepancy is due to the difference in their ages; we explain the interaction between time, species number and disparity in greater detail in the text.